

Supplementary Information for:

Day-to-day associations between testosterone, sexual desire, and courtship efforts in young men

Tikal Catena, Blair T. Crewther, Adar B. Eisenbruch, Rachel L. Grillot, Dario Maestripieri,
James R. Roney

Proceedings of the Royal Society of London B

DOI: doi.org/10.1098/rspb.2024.1508

Table of Contents:

S1 – Supplementary Analyses

S1.1 – Power analyses

S1.2 – Differences in key variables by relationship status

S1.3 – Further exploration of variability in testosterone-desire relationships

S1.4 – Testosterone-desire relationships in single vs. partnered participants

S1.5 – Testosterone-courtship effort relationships – supplementary analyses

S1.6 – Testing the dual-hormone hypothesis

S2 – Robustness checks

S2.1 - Sexual desire models

S2.2 Courtship effort models

Overview:

This document contains supplementary documents S1 and S2 (outlined in the table of contents above). S1 describes the results of supplementary analyses that are referenced throughout the main text. S2 describes a series of robustness checks, in which we examine how reversing exclusion criteria and variable transformations affect the results of analyses presented in the main text.

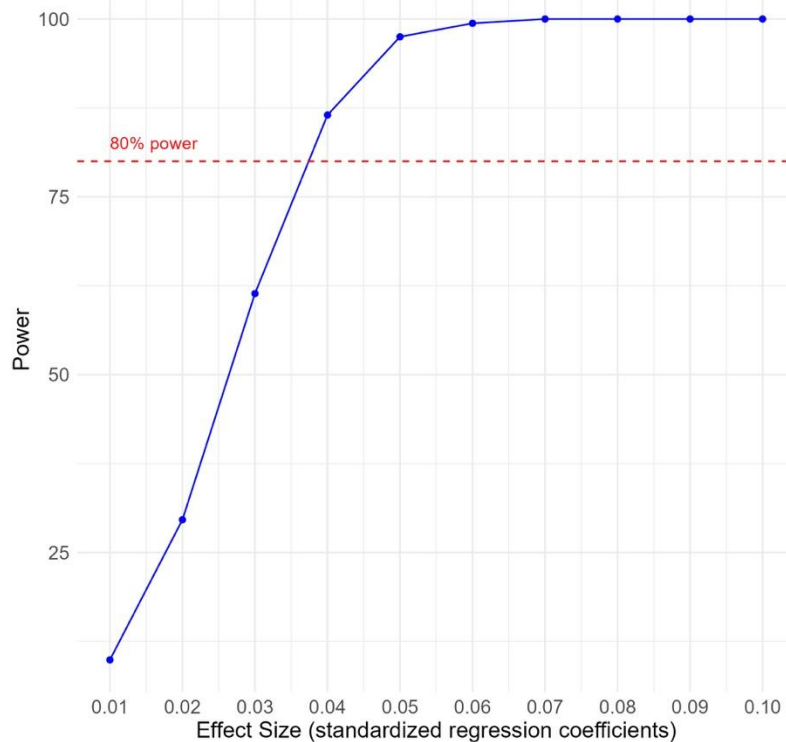
Documents S3, S4, and S5 can be accessed in our OSF repository (https://osf.io/bzvdf/?view_only=1e50e7c980834a779d032a22b1752cfc). S3 contains the R code and raw model outputs for analyses referenced in S1 and S2, with the exception of the code for our power analyses. S4 contains the code for the power analyses. S5 contains code and raw model outputs for analyses presented in the main text.

S1 – Supplementary Analyses

S1.1 – Power analyses

To verify that we had the statistical power to detect small effects, we conducted a series of power analyses using the `simr` R package. We simulated datasets reflecting the within and between-subject variance in sexual desire observed in our mixed-effects model reported in the main text, along with varying hypothetical effect sizes (beta coefficients ranging from .01 to .1) that could have existed between observed testosterone concentrations and sexual desire. Using `simr`, at 1000 simulations per effect size, we measured the likelihood of our final sample size (41 subjects, with 22 observations each) being sufficient to detect the simulated effect. The results can be seen in Figure S1, with full code reported in S4.1.

Figure S1 – Power for detecting increasing effect sizes with our final sample size



As can be seen above, we had sufficient power (over 80%) to detect effect sizes as small as 0.04, and abundant power to detect a small effect size of 0.1. An equivalent analysis was carried out to assess our power to detect a relationship between testosterone and courtship effort and found similar results (over 80% power to detect effect sizes as small as 0.04). The code used for these calculations is provided in S4.2.

We also conducted power analyses to assess our power to detect level-2 effects of mean hormone concentrations. These were conducted using a more direct approach in base R, simulating our power to detect varying effect sizes with 41 independent samples (reflecting the number of hormone means we measured). We found that we had 80% power to detect between-subject effects of 0.6 or higher in our sample (the code for these calculations can be found in S4.3).

S1.2 – Differences in key variables by relationship status

We examined differences between single and partnered participants in key variables using mixed-effects linear models with subject-level random intercepts. We also calculated marginal means for each group based on the linear models, using the ‘emmeans’ package in R. Results are reported below in Table S1. Single participants had higher testosterone and cortisol concentrations than partnered participants, and there was a trend toward higher desire in partnered participants and higher courtship effort in single participants.

Table S1 – Differences in key variables between single and partnered participants.

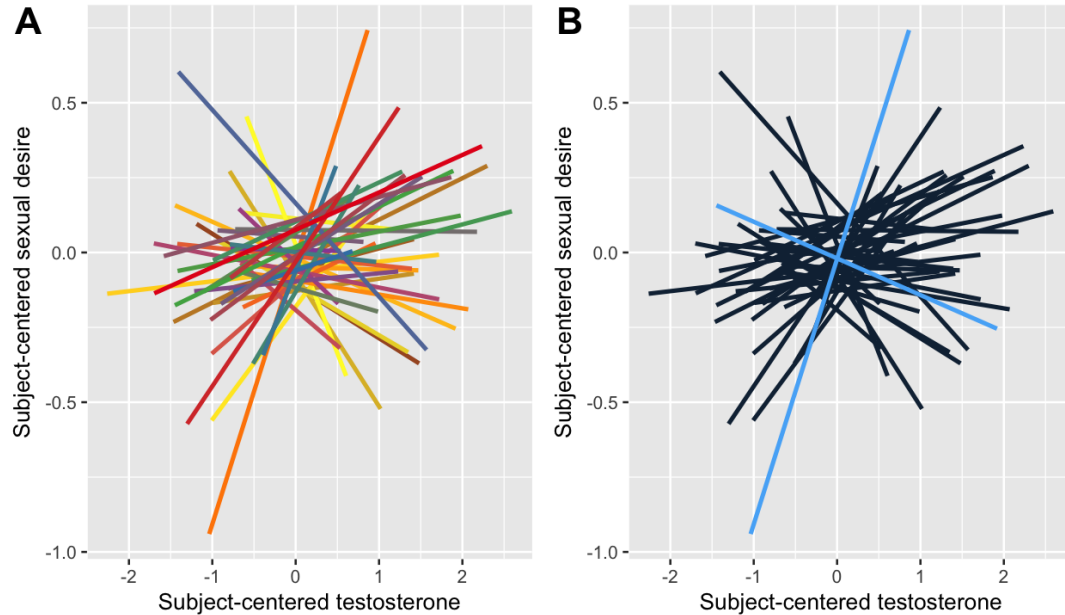
Dependent variable	t-score	df	p-value	Single participant mean z-score	Partnered participant mean z-score
Sexual Desire	1.661	38.939	0.105	-0.182	0.264
Courtship Effort	-1.625	39.083	0.112	0.121	-0.277
Testosterone	-3.258	38.830	0.002	0.277	-0.462
Cortisol	-2.849	38.265	0.007	0.164	-0.305

S1.3 – Further exploration of variability in testosterone-desire relationships

We examined in further detail what individual differences in testosterone-desire relationships may be identifiable in our sample. First, we calculated regression coefficients and *p*-values for each subject. Only two participants showed statistically significant testosterone-desire correlations; one positive ($r = .595, p = .0095$) and one negative ($r = -.534, p = .0154$).

Overall, this indicates a much lower degree of variability in testosterone-desire slopes than might at first be apparent from the data, which may be highlighted by the comparison between the spaghetti plot presented in the manuscript (reproduced below as Fig. S2A) and the same figure with only the significant correlation lines colored (Fig. S2B). Although there seem to be many individuals with strong testosterone-desire relationships in Figure S2A, most of these correlations are not significantly different from zero.

Figure S2 – Subject level testosterone-desire slopes



To further assess whether the variability in testosterone-desire regression slopes in our sample was consistent with random noise, we generated three simulated datasets of the same size as our real dataset. Using the base ‘rnorm’ function in R, random (uncorrelated) testosterone and sexual desire values were generated for 20 data points nested within 41 participant ID numbers. This process was repeated three times with different randomization seeds to generate replicable results (R code and random seeds can be found in S3.1). Subject level slopes for each dataset were then plotted in the same manner as in Fig. 1, with spaghetti plots generated from each of the respective data sets presented in Figures S3, S4, and S5 presented below. Panel B in each of the figures highlights the statistically significant correlations.

Figure S3 – Subject-level slopes in randomly generated data (random dataset 1)

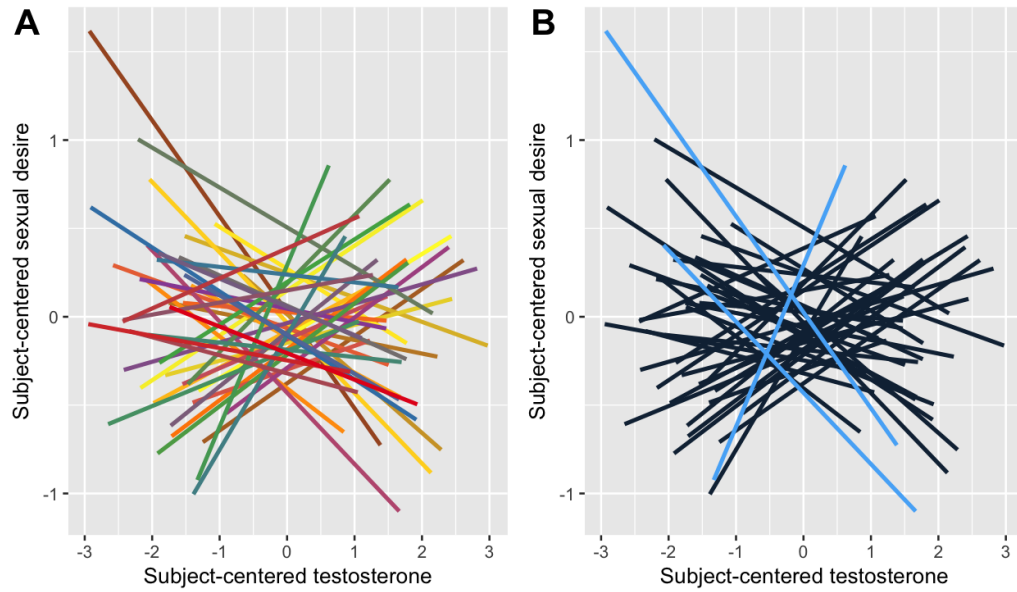


Figure S4 – Subject-level slopes in randomly generated data (random dataset 2)

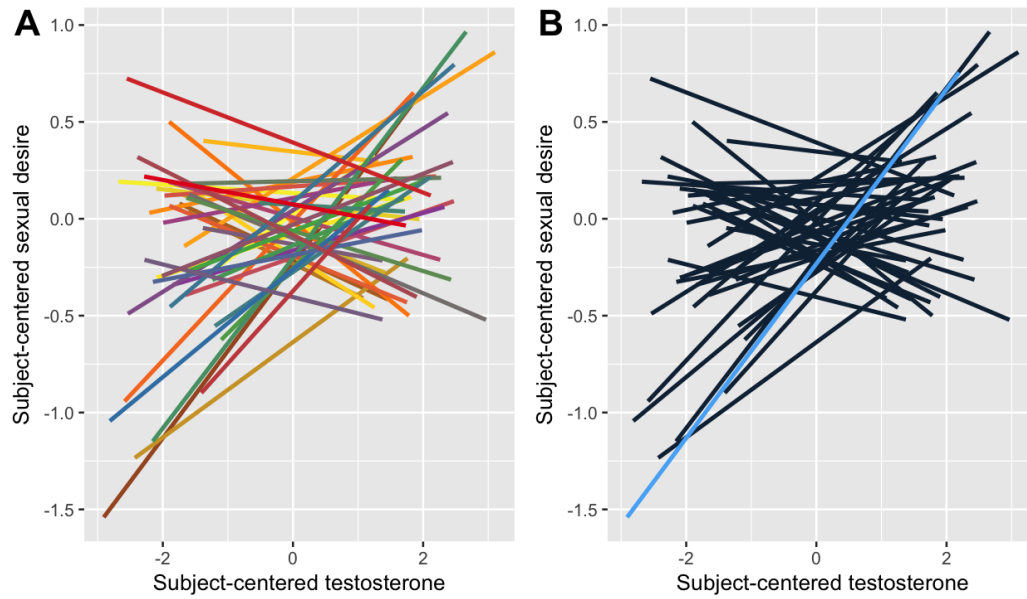
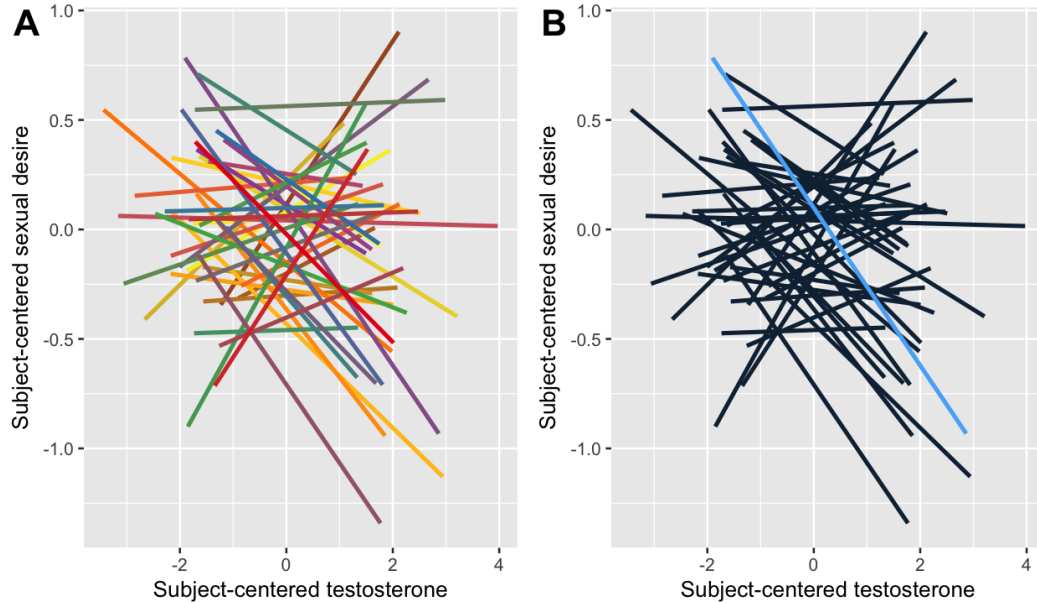
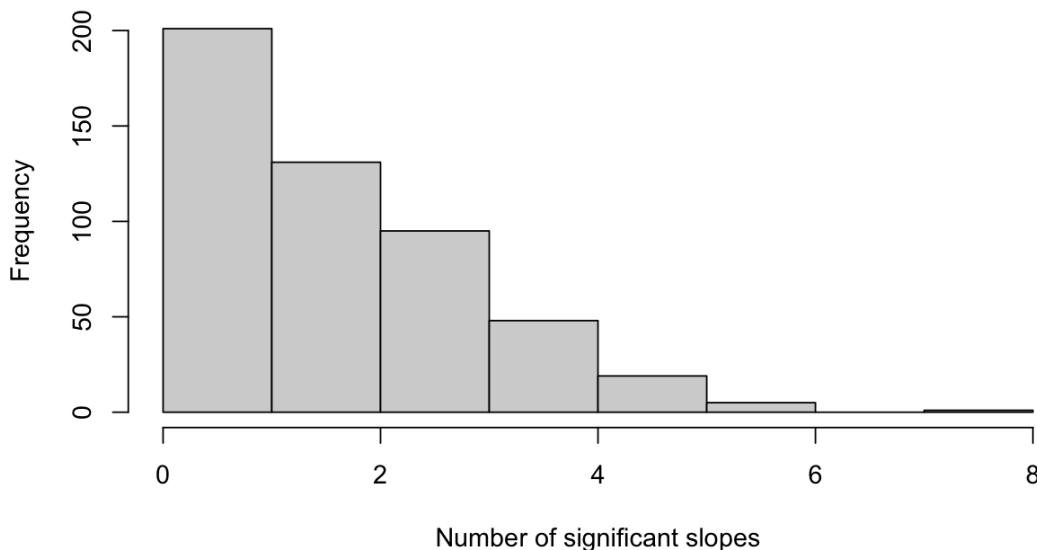


Figure S5 –Subject-level slopes in randomly generated data (random dataset 3)



As can be seen from comparing Fig. 1B in the main text to Figs. S3 to S5, the distribution of subject-level slopes observed in our data is similar to those generated from completely uncorrelated data. Furthermore, to establish a baseline of what number of significant correlations we should expect from uncorrelated data, a bootstrapped analysis was carried out, simulating 500 datasets with random testosterone and desire values nested within 41 participants. The overall distribution of significant correlations found in each dataset can be seen in Figure S6. The mean number of significant correlations across simulated datasets was 1.955, comparable to the two significant correlations observed in our data.

Figure S6 – Statistically significant testosterone-desire slopes in 500 simulated datasets.



S1.4 – Testosterone-desire relationships in single vs. partnered participants.

S1.4.1 – Linear mixed regression models

To assess whether effects of daily testosterone differed by relationship status, we split the dataset between single and partnered participants, and ran separate mixed-effects linear regression models for each subsample. As in the main text, models predicted daily sexual desire from daily testosterone and mean testosterone, with random subject-level intercepts. There was no significant effect of daily testosterone for single ($\beta = 0.026$, $df = 443$, $p = 0.433$, 95% CI [-0.049, 0.116]) or partnered ($\beta = 0.037$, $df = 228$, $p = 0.246$, 95% CI [-0.044, 0.173]) participants.

S1.4.2 – CT Models

Table S2 shows the CT model fit for testosterone and sexual desire auto-effects, cross-effects, and diffusion covariance for all participants. As can be seen in Table S2, the diffusion covariance between testosterone and sexual desire was marginally significant. Significant diffusion covariance between testosterone and sexual desire indicates a positive association between the residual error variance (as in regular regression) in the testosterone and sexual desire processes. This may reflect shared causes for within-subject variation between testosterone and sexual desire on a given day.

Tables S3 and S4 report the model fits for the same models when splitting the data between single and partnered participants (respectively). When dividing the data in this way, the negative lagged effect of testosterone on sexual desire remained only for partnered participants. The diffusion covariance effect, on the other hand, converged as a significant effect for both subsamples.

Table S2. Continuous-time parameter estimates for the testosterone-sexual desire bivariate model in all participants.

Parameter Type	Parameter	Estimate	SE	p-value	95% CI
Auto-effect	Testosterone	-1.525	0.146	< 0.001	[-2.012, -1.038]
	Sexual Desire	-1.137	0.128	< 0.001	[-1.623, -1.123]
Cross-effect	Testosterone → Sexual Desire	-0.460	0.184	0.012	[-0.821, -0.099]
	Sexual Desire → Testosterone	0.085	0.153	0.58	[-0.215, 0.385]
Diffusion Covariance	Testosterone ↔ Sexual Desire	0.472	0.241	0.0504	[-0.001, 0.944]

Table S3. Continuous-time parameter estimates for the testosterone-sexual desire bivariate model for single participants

Parameter Type	Parameter	Estimate	SE	p-value	95% CI
Auto-effect	Testosterone	-1.373	0.243	< 0.001	[-1.849, -0.897]
	Libido	-1.246	0.130	< 0.001	[-1.501, -0.991]
Cross-effect	Testosterone → Sexual Desire	-0.218	0.206	0.292	[-0.622, 0.186]
	Sexual Desire → Testosterone	0.149	0.166	0.369	[-0.176, 0.474]
Diffusion Covariance	Testosterone ↔ Sexual Desire	2.381	0.232	< 0.001	[1.926, 2.836]

Table S4. Continuous-time parameter estimates for the testosterone-sexual desire bivariate model for partnered participants

Parameter Type	Parameter	Estimate	SE	p-value	95% CI
Auto-effect	Testosterone	-2.265	0.802	0.005	[-3.837, -0.693]
	Sexual Desire	-1.781	0.437	< 0.001	[-2.638, -0.924]
Cross-effect	Testosterone → Sexual Desire	-1.243	0.540	0.021	[-2.301, -0.185]
	Sexual Desire → Testosterone	-0.070	0.475	0.883	[-1.001, 0.861]
Diffusion Covariance	Testosterone ↔ Sexual Desire	3.541	0.855	< 0.001	[1.865, 5.217]

S1.5 – Testosterone-courtship effort relationships – supplementary analyses

S1.5.1 – Linear mixed regression models for single and partnered participants

As in S4.1.1, we tested whether the relationship between testosterone and courtship effort varied when the dataset was split between single and partnered participants. These models predicted daily courtship effort from daily testosterone and mean testosterone, with random subject-level intercepts. We found no significant effects of daily testosterone in single ($\beta = 0.053$, $df = 443$, $p = 0.119$, 95% CI [-0.02, 0.173]) or partnered ($\beta = 0.006$, $df = 228$, $p = 0.889$, 95% CI [-0.095, 0.11]) participants.

S1.5.2 – CT model parameter estimates

Table S5 shows the parameter estimates for the continuous-time model for testosterone and courtship effort reported in the main text. There were significant auto-effects for both testosterone and courtship effort, no cross-effects, and significant diffusion covariance. As mentioned above, diffusion covariance represents the within-subject correlation in the random changes in each time series and thus, reflects the extent to which these processes may share common causes.

Table S5. Continuous-time parameter estimates for the testosterone-courtship effort bivariate model

Parameter Type	Parameter	Estimate	SE	p-value	95% CI
Auto-effect	Testosterone	-1.667	0.350	< 0.001	[-2.353, -0.981]
	Courtship Effort	-1.955	0.270	< 0.001	[-2.484, -1.426]
Cross-effect	Testosterone → Courtship Effort	-0.602	0.363	0.098	[-1.313, 0.109]
	Courtship Effort → Testosterone	-0.307	0.283	0.278	[-0.862, 0.248]
Diffusion Covariance	Testosterone ↔ Courtship Effort	3.701	0.518	< 0.001	[2.686, 4.716]

S1.6 – Testing the dual-hormone hypothesis

Below we report summarized results of analyses predicting outcomes of interest from cortisol concentrations. The R code and model results can be seen in full in S3.5. Cortisol measures were treated the same as testosterone measures. First, such measures were converted to residuals from time since waking. These residuals were then grand-mean standardized and centered within-subjects.

To test the predictions of the dual-hormone hypothesis, we ran a mixed-effects model predicting sexual desire from the interaction between testosterone and cortisol. To match the structure of the main text models, we also included subjects' mean testosterone, mean cortisol, and relationship status as predictors. The model converged with random slopes only for the effect of same-day cortisol. There were no significant effects (see Table S6; for R code and full model results, see S3.5.2).

In addition to examining sexual desire, we also ran a mixed-effects model predicting courtship effort from the interaction between testosterone and cortisol. This, too, revealed no significant relationships (all p s > 0.05; for full analyses, see S3.5.2).

We also examined mixed-effects linear models including only cortisol measures and relationship status, with random subject-level intercepts and slopes, predicting daily sexual desire. There was no significant effect of cortisol on sexual desire ($\beta = 0.003$, $df = 35$, $p = 0.75$). A mixed-effects model with random subject-level intercepts was also run predicting courtship effort from cortisol concentrations, and revealed no significant fixed effect ($\beta = 0.034$, $df = 671$, $p = 0.24$). Full R code and model outputs can be found in S3.5.1.

Table S6. Fixed and random effects for the mixed-effect model predicting same-day sexual desire from testosterone, cortisol, and their interaction.

Fixed Effects						
Item	Estimate	Std. Error	df	t	P-value	95% CI
(Intercept)	-0.119	0.166	37.011	-0.717	0.478	[-0.435, 0.198]
Mean cortisol	0.124	0.265	37.082	0.466	0.644	[-0.386, 0.633]
Same-day cortisol	0.003	0.031	37.050	0.104	0.918	[-0.058, 0.064]
Testosterone x cortisol	0.008	0.038	449.026	0.218	0.828	[-0.066, 0.083]
Relationship status	0.266	0.314	36.941	0.848	0.402	[-0.333, 0.864]
Mean testosterone	-0.313	0.196	36.935	-1.596	0.119	[-0.697, 0.072]
Same-day testosterone	0.037	0.035	656.418	1.060	0.289	[-0.032, 0.106]
Random Effects						
Item	Estimate	Std. Error	Wald's Z	p-value	95% CI	
(Intercept)	.638	.153	4.165	< .001	[0.338, 0.938]	
Same-day cortisol	.005	.019	.279	0.781	[-0.032, 0.043]	

S2 – Robustness checks (complete code and model statistics reported in S3)

S2.1 - Sexual desire models

S2.1.1 – Testing the robustness of main text models

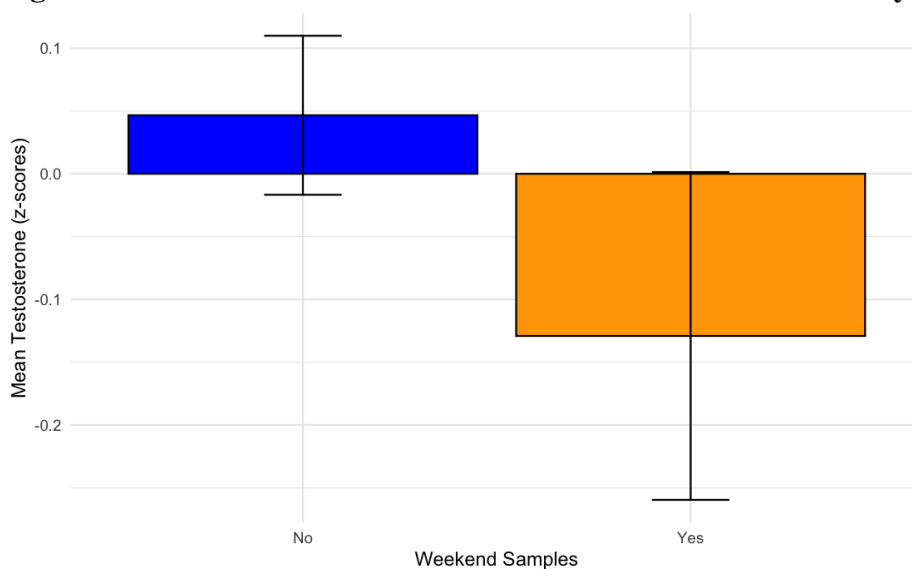
Two modifications were made to raw testosterone measures used in the main text models. First, testosterone concentrations were replaced with residuals extracted from time-since-waking analyses (see main text section 2.d). Second, outlier testosterone concentrations more than 3 standard deviations from the population mean were excluded from analyses.

To check the robustness of our findings, we reproduced the main-text models predicting sexual desire with three alternative testosterone measures: 1) residuals from time since waking, with outliers left in; 2) raw testosterone values, with outliers left in; 3) raw testosterone values, with outliers excluded. In all three analyses, testosterone was not significantly associated with sexual desire (see S3.6.1.1 for full analyses). As in the main text models, these models also included relationship status and mean testosterone as subject-level predictors.

S2.1.2 – Incorporating weekend samples

Converting testosterone measures to residuals from time since waking required the exclusion of weekend samples; as discussed in the main text, these samples were missing associated time since waking measures due to a clerical error. The decision to exclude weekend samples was further justified by systematic differences in testosterone concentrations found between weekend and weekday samples. Specifically, a mixed-effects model was run predicting testosterone from a weekend-sample binary variable, with subject-level random slopes and intercepts. This revealed significantly lower testosterone concentrations for weekend samples relative to weekday samples, as seen in Figure S7 ($\beta = -0.076$, $df = 39$, $p < 0.05$).

Figure S7 – Differences in testosterone measures between weekday and weekend samples



One possible reason for this pattern is that saliva samples on weekends were perhaps collected at a later time than weekday samples, and therefore often referenced a later sampling time from participants' natural circadian rhythm. If this were the case, it would imply that weekend testosterone measures understated subjects' testosterone concentrations on those days relative to the weekday measures. Alternatively, testosterone concentrations may have truly been lower on weekends due to other unknown factors.

Additional issues were presented by the fact that our outcome measures were higher on weekends, based on mixed-effects models comparing weekdays and weekends (for R code and full model results see S3.6). Weekend responses had significantly higher reported sexual desire ($\beta = 0.041$, $df = 37$, $p < .05$) and marginally higher courtship effort ($\beta = 0.039$, $df = 37$, $p = 0.066$). These patterns further complicate the use of weekend hormone values given that testosterone concentrations from weekend samples may be systematically underestimated in the models.

With this in mind, we still carried out two further analyses: 1) including all raw testosterone measures (including weekend samples) as predictors; 2) including all raw testosterone measures but excluding outliers (more than 3 SDs from the mean). As with the previous section, neither of these analyses revealed a significant association between testosterone and sexual desire (see S3.6.1.2 for full analyses).

S2.2 Courtship effort models

S2.2.1 – Testing the robustness of main text models

We carried out the same procedure as in S2.1.1 to test the robustness of the main text courtship effort models, with the same variations on weekday testosterone measures. The zero-order relationship between testosterone and courtship effort remained non-significant in all analyses (see S3.6.2 for full analyses). In the models predicting single men's courtship effort from the interaction between testosterone and DSI, testosterone remained a significant predictor of courtship effort (Table S7).

Table S7 – Relationship between testosterone and courtship effort in the interaction models (single men)

Testosterone variable	Relationship between testosterone and courtship effort
Residuals from time since waking, outliers left in	$\beta = 0.125$, $df = 11$, $p < 0.01$, 95% CI [0.069, 0.262]
Raw testosterone, outliers left in	$\beta = 0.102$, $df = 16$, $p < 0.05$, 95% CI [0.035, 0.224]
Raw testosterone, outliers excluded	$\beta = 0.095$, $df = 12$, $p < 0.05$, 95% CI [0.048, 0.199]

As in the main text, we ran follow-up analyses for the above models examining the relationship between testosterone and courtship effort on DSI days only. These models converged with subject-level random intercepts and random slopes for the daily testosterone effect. Results showed that the significant effect of daily testosterone remained significant across all models (Table S8).

Table S8– Relationship between testosterone and courtship effort on DSI days (single men)

Testosterone variable	Relationship between testosterone and courtship effort
Residuals from time since waking, outliers left in	$\beta = 0.177$, $df = 7$, $p < 0.05$, 95% CI [0.065, 0.435]
Raw testosterone, outliers left in	$\beta = 0.151$, $df = 10$, $p < 0.05$, 95% CI [0.025, 0.381]
Raw testosterone, outliers excluded	$\beta = 0.141$, $df = 6$, $p < 0.05$, 95% CI [0.018, 0.348]

S2.2.2 – Incorporating weekend samples

As in section S2.1.2, we also repeated the above analyses incorporating weekend samples, creating two more alternative models: 1) using raw testosterone values from all days, including outliers; 2) using raw testosterone values from all days, but excluding outliers. When using these testosterone values, no effects were significant (see S3.6.2.2 for zero-order analyses; see S3.6.3.2 for interaction analyses).